Published in final edited form as:

Med Image Comput Assist Interv. 2016 October ; 9901: 247–255. doi:
10.1007/978-3-319-46723-8_29.

Automatic Cystocele Severity Grading in Ultrasound by Spatio-Temporal Regression

Dong Ni¹, Xing Ji¹, Yaozong Gao², Jie-Zhi Cheng¹, Huifang Wang³, Jing Qin⁴, Baiying Lei¹, Tianfu Wang¹, Guorong Wu², and Dinggang Shen²

¹National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University, Shenzhen, China

²Department of Radiology and BRIC, UNC at Chapel Hill, Chapel Hill, NC 27599, USA

³Department of Ultrasound, Shenzhen Second Peoples Hospital, Shenzhen, China

⁴School of Nursing, Centre for Smart Health, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Abstract

Cystocele is a common disease in woman. Accurate assessment of cystocele severity is very important for treatment options. The transperineal ultrasound (US) has recently emerged as an alternative tool for cystocele grading. The cystocele severity is usually evaluated with the manual measurement of the maximal descent of the bladder (MDB) relative to the symphysis pubis (SP) during Valsalva maneuver. However, this process is time-consuming and operator-dependent. In this study, we propose an automatic scheme for cystocele grading from transperineal US video. A two-layer spatio-temporal regression model is proposed to identify the middle axis and lower tip of the SP, and segment the bladder, which are essential tasks for the measurement of the MDB. Both appearance and context features are extracted in the spatio-temporal domain to help the anatomy detection. Experimental results on 85 transperineal US videos show that our method significantly outperforms the state-of-the-art regression method.

Keywords

Ultrasound; Regression; Spatio-temporal; Cystocele

1 Introduction

Cystocele is a common disease in woman that occurs when bladder bulges into vagina due to defects in pelvic support. The accurate assessment of cystocele severity is very important for treatment options, which can be no treatment for a mild case or surgery for a serious case. Pelvic Organ Prolapse Quantification system (POP-Q) is widely used for cystocele diagnosis [1]. This evaluation system involves many complicated procedures and may be clinically inefficient [2]. Recently, the transperineal ultrasound (US) has emerged as a new and

effective tool for cystocele diagnosis for its advantages of no radiation exposure, minimal discomfort, cost-effectiveness and real-time imaging capability [3]. Generally, the US examination for cystocele includes four steps [4] (Fig. 1). First, a radiologist steadily holds the US probe on the patient when asking the patient to perform Valsalva maneuver. Then, an image frame containing the maximal descent of the bladder (MDB) relative to the symphysis pubis (SP) is manually selected from US video. Next, the MDB is manually measured as the distance from the lowest point of the bladder to the reference line. With the measured MDB, the degree of cystocele severity can be further graded into normal, mild, moderate, and severe. In these steps, frame selection and manual measurements are time-consuming and experience-dependent, which often leads to significant inter-observer grading variations [5]. Therefore automatic methods for cystocele grading may help to improve diagnostic efficiency and decrease inter-observer variability.

As shown in Fig. 1, the identification of the middle axis and lower tip of SP and bladder segmentation in US images deem to be necessary tasks for severity grading. However, these tasks are very challenging. First, due to the vagueness in US images, the localization of SP and its lower tip is very difficult, even for a senior radiologist. Second, the missing or weak boundaries of the bladder resulted from acoustic attenuation, speckles and shadows make the segmentation task difficult. Third, the image appearance, geometry and shape of anatomies vary significantly in the US image series of Valsalva maneuver, because of forced exhalation. They also vary significantly from subject to subject. These large variations will then impose additional difficulty for our automation goal.

In this study, a novel spatio-temporal regression model is proposed to address the three challenging issues for the automatic analysis of transperineal US video and cystocele grading. The technical contributions of this work are summarized as follows. First, to our knowledge, this is the first study that performs the computerized grading of cystocele severity with the transperineal US. Second, we propose a two-layer spatio-temporal regression model for context-aware detection of anatomical structures at all time points jointly. In our proposed model, both appearance and context features are extracted in the spatio-temporal domain to impose temporal consistency along the temporal displacement maps, thus the detection results can help each other to alleviate the ambiguity and refine structure localization.

2 Method

For the automatic grading of cystocele severity, we first train the two-layer spatio-temporal regression models for the identification of the middle axis and lower tip of SP and segmentation of bladder in US images. With the trained models, the descending of the bladder relative to the SP was measured in all image frames of a Valsalva maneuver US video. The MDB can then be sought from the estimated distance measurements over all US frames for cystocele grading.

2.1 The Proposed Spatio-Temporal Regression Model

Random forest [6] is an ensemble learning technique with good generalization capability [7]. This technique has been successfully applied in many medical image analysis tasks, e.g.,

landmark detection, organ segmentation and localization [8–10], etc. Here we employ the random forest to train the two-layer spatio-temporal regression models for the detection of target structures in US videos.

To build a random forest, multiple decision trees are constructed by randomly sampling the training data and features for each tree to avoid over-fitting. The final regression result, $P(d^s/\mathbf{v})$, can then be reached by averaging the predictions of T trees, $p_i(d^s/\mathbf{v})$, as:

$$P(d^s(\mathbf{x})|\mathbf{v}(\mathbf{x})) = \frac{1}{T} \sum_{i=1}^T p_i(d^s(\mathbf{x})|\mathbf{v}(\mathbf{x})) \quad (1)$$

where $\mathbf{x}$ is the image pixel, $\mathbf{v}$ is the feature vector and d^s is the distance of $\mathbf{x}$ to the target structure s , and $s \in \{l, t, b\}$. The target structures l , t and b represent the middle axis and lower tip of the SP and the bladder, respectively.

As shown in Fig. 2, we train one regression forest for each target structure s , to learn its specific non-linear mapping from each pixel's local appearance and geometry to its 2D displacement vector towards the specific structure. Specifically, the first layer is designed to provide the initial displacement field for each time point by using the appearance and coordinates features from neighboring US images, while the second layer is designed to refine the detection result in spatio-temporal domain (a 2D+t neighborhood) by using contexture features from the results in the first layer.

First-Layer Regression—The SP appears like a large bright ridge with two dark valleys around in US images (see Fig. 1), whereas a bladder is depicted with hypoechogenicity in sonography for its fluid content. Accordingly, contrast features shall be informative and helpful for modeling of these structures. Furthermore, the correlation between neighboring US frames can be utilized as temporal consistency for displacement field. In this regard, we compute randomized Haar-like features [11] of different scales in spatio-temporal domain to describe the intensity patterns and the contrastness of target structures, as well as to boost anatomy detection at current time point with additional temporal cues from previous and next time points. Meanwhile, we also use normalized coordinate as input features. With these features, we train the regression forest to seek a reliable nonlinear mapping that tells the displacement vector of a pixel to the target structures of the middle axis and lower tip of the SP and the bladder, denoted as d^l , d^t , and d^b , respectively. The definitions of the displacement maps for the three target structures can be seen in Fig. 3.

Second-Layer Regression—We first use the above trained first-layer regression forest to estimate an initial displacement map at each time point. Thus, for each image pixel, we have not only appearance features but also additional high-level context feature [12] from the initial displacement map at current time point and along all other displacement maps at other time points. All these features are used to train the second-layer regression forest jointly. Specifically, our context features are calculated again by Haar-like features from local patches in the displacement maps. Two types of context features are extracted: (1) *Within-time-point context features* refer to the Haar-like features extracted within the

displacement map of each structure. These features are informative in providing the estimated structure locations from nearby pixels, and can be used to spatially regularize the whole displacement of each structure. (2) *Across-time-point context features* refer to the Haar-like features extracted from the displacement maps of the same structure at other time points. These features encode the temporal relationship along time, i.e., the trajectory of structure. Thus, the use of across-time-point context features can effectively impose temporal consistency on the displacement field. With the augmented feature vector, we perform the random forest regression again to approach the target distance spaces of d^l , d^m , d^b .

2.2 Cystocele Severity Grading

With the two-layer random forest regressors, the middle axis and lower tip of the SP and the bladder contour can be inferred for the MDB measurement and severity grading. We first generate the displacement maps of the three target structures from the testing sonography. The voting maps is then obtained for the lower tip and middle axis of the SP by adopting the voting strategy in [8] on the corresponding displacement maps. Next, the lower tip of the SP can be identified by searching the most votes in its voting map. Then, the delineation of the middle axis of the SP can be realized by seeking the line that originates from lower tip with maximal average voting responses. For the bladder segmentation in the testing sonography, the bladder boundary can be simply attained by finding the zero level set on its displacement feature map. Once the three target anatomies are defined, we calculate the MDB from the consecutive US images (Fig. 1). Then, we categorize the severity degree of cystocele into normal, mild, moderate, and severe by adopting the thresholds of the MDB recommended in [13].

3 Experimental Results

Materials

We acquired 170 US videos from 170 women with ages ranging from 20 to 41. Each video lasts approximately 10 s and contains around 400 frames. The data is randomly split into 85 and 85 videos for the training and testing, respectively. All videos were acquired using a Mindray DC8 US scanner with local IRB approvals. To support the training of regression models, one graduate student was recruited to prepare the necessary annotation on each training image. The annotated training data were further reviewed by a senior radiologist with experience on medical US over 15 years to assure correctness. The number of neighboring frames for extracting spatio-temporal features was 30 and other parameters were set according to [11]. To evaluate the performance of our system and the inter-observer variation, three radiologists with US imaging experience of more than 3 years were invited to annotate the middle axis and lower tip of SP on each testing image. Each radiologist was also asked to measure the bladder descent on each testing image and give the cystocele severity grades of all patients. The bladder boundaries were not annotated in the testing data as the boundary drawing task is very costly.

Intermediate Results

We first evaluate the performance on the identification of the middle axis and lower tip of SP. Figure 5 shows the comparison of the performance of our automatic system on four typical cases with the three sets of manual annotations. It can be found that there exists significant variation of SP and bladder in terms of shape, geometry and appearance. Our method can generate the reasonably good intermediate results by comparing to the manual definitions. We further evaluate the MDB performance by comparing the accuracies of the MDBs from spatio-temporal regression model (2D+t) and the regression model without temporal cue (2D) [11]. The means and standard deviations of absolute MDB differences of the proposed method and three radiologists (namely E1, E2 and E2) are 3.02 ± 2.74 mm, 3.01 ± 2.59 mm and 3.00 ± 2.91 mm, respectively, whereas the differences between the MDBs of 2D regression [11] and three radiologists are 3.92 ± 3.04 mm, 4.68 ± 3.19 mm and 4.78 ± 3.50 mm, respectively. The p-values (two-sample, two-tailed *t*-test) between two automatic methods w.r.t. three radiologists are 0.0287, 6.8538e-04 and 9.2093e-04, respectively. It can then be concluded our spatio-temporal model is significantly better than the regression method without temporal cue. The boxplots of the MDB measurements by two methods are also shown in Fig. 4.

Accuracy of Cystocele Severity Grading

Here we show the clinical applicability by comparing final grading results of two automatic methods. The Cohens kappa statistics is used to evaluate the grading agreement between the radiologists and the computerized methods. As illustrated in Table 1, the overall grading accuracies to three radiologists by our proposed method (2D+t) are all higher than 80 %. The grading results by our method are significantly better than the 2D regression method [11]. The Kappa values shown in Table 1 further indicate that our method can achieve significantly better agreement with the radiologists than the 2D regression method. It can then be suggested the incorporation of temporal appearance and context features into the random forest regression is effective. We further calculate the Kappa values of the manual grading results by three radiologists to compare the agreement between the radiologist to the computer as well as the inter-radiologist agreement. The Kappa values of radiologists are 0.65 (E1 vs. E2), 0.55 (E1 vs. E3) and 0.87 (E2 vs. E3), respectively. It can be suggested that the grading agreements between the computer and radiologists are relatively stable, comparing to inter-radiologist agreement. In particular, the grading results between the radiologist 1 and other radiologists are relatively less consistent.

4 Conclusions

This paper develops the first automatic solution for grading cystocele severity in the transperineal US videos. A novel spatio-temporal regression model is proposed to introduce temporal consistency for displacement field estimation. Both appearance and context features in spatio-temporal domain can boost the anatomy detection performance in US images. The experimental results suggest that our method significantly outperforms the 2D regression method in terms of intermediate distance measurement and final severity grading. The developed system is robust and has potential in clinical applicability.

Acknowledgments

This work was supported by the National Natural Science Funds of China (Nos. 61501305, 61571304, and 81571758), the Shenzhen Basic Research Project (Nos. JCYJ20150525092940982 and JCYJ20140509172609164), and the Natural Science Foundation of SZU (No. 2016089).

References

1. Persu C, Chapple C, Cauni V, Gutue S, Geavlete P. pelvic organ prolapse quantification system (POP-Q)-a new era in pelvic prolapse staging. *J Med Life*. 2011; 4(1):75. [PubMed: 21505577]
2. Lee U, Raz S. Emerging concepts for pelvic organ prolapse surgery: what is cure? *Cur Urol Rep*. 2011; 12(1):62–67.
3. Santoro G, Wiczeorek A, Dietz H, Mellgren A, Sultan A, Shobeiri S, Stankiewicz A, Bartram C. State of the art: an integrated approach to pelvic floor ultrasonography. *Ultrasound Obstet Gynecol*. 2011; 37(4):381–396. [PubMed: 20814874]
4. Chan L, Tse V, Stewart P. Pelvic floor ultrasound. 2015
5. Thyer I, Shek C, Dietz H. New imaging method for assessing pelvic floor biomechanics. *Ultrasound Obstet Gynecol*. 2008; 31(2):201–205. [PubMed: 18254157]
6. Breiman L. Random forests. *Mach Learn*. 2001; 45(1):5–32.
7. Verikas A, Gelzinis A, Bacauskiene M. Mining data with random forests: a survey and results of new tests. *Pattern Recogn*. 2011; 44(2):330–349.
8. Gao, Y., Shen, D. Context-aware anatomical landmark detection: application to deformable model initialization in prostate CT images. In: Wu, G.Zhang, D., Zhou, L., editors. *MLMI 2014. LNCS*. Vol. 8679. Springer; Heidelberg: 2014. p. 165-173.
9. Richmond, D., Kainmueller, D., Glocker, B., Rother, C., Myers, G. Uncertainty-driven forest predictors for vertebra localization and segmentation. In: Navab, N.Hornegger, J.Wells, WM., Frangi, AF., editors. *MICCAI 2015. LNCS*. Vol. 9349. Springer; Heidelberg: 2015. p. 653-660.
10. Zhou, SK., Comaniciu, D. Shape regression machine. In: Karssemeijer, N., Lelieveldt, B., editors. *IPMI 2007. LNCS*. Vol. 4584. Springer; Heidelberg: 2007. p. 13-25.
11. Shao Y, Gao Y, Wang Q, Yang X, Shen D. Locally-constrained boundary regression for segmentation of prostate and rectum in the planning CT images. *Med Image Anal*. 2015; 26(1): 345–356. [PubMed: 26439938]
12. Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2010; 32(10):1744–1757. [PubMed: 20724753]
13. Wang H, Chen H, Zhe R, Xu F, Chen Q, Liu Y, Guo J, Shiya W. Correlation between anterior compartment prolapse assessments by transperineal ultrasonography and pelvic organ prolapse quantification. *Chin J Ultrason*. 2013; 22(8):684–687.

Fig. 1.

Illustration of the MDB measurement. Several US snapshots acquired during Valsalva maneuver are listed in the upper row. The lower row shows the process of MDB measurement. The MDB (in green, sub-figure (c)) is measured as the distance between the reference line (RL, in blue) and the lowest point of the bladder (BL) relative to the RL. The RL originates from the lower tip of the SP and its direction is 135 degree clockwise from the middle axis of the SP.

Fig. 2.
The flowchart of proposed two-layer spatio-temporal regression model.

Fig. 3.
The distance definition with respect to three target structures.

Fig. 4.
Boxplots of the MDB distributions.

Fig. 5.

Comparison of measurements by our method (in red) and 3 radiologists (in yellow, green and purple). The severities are graded into normal, mild, moderate and severe from the top to the bottom videos. The sub-figure marked by red box contains the maximal descent of the bladder from the SP.

Table 1

Overall grading accuracy and Kappa statistics.

		Auto vs. E1	Auto vs. E2	Auto vs. E3
Grading accuracy	2D	78.82 %	74.12 %	75.29 %
	2D+t	87.06 %	81.18 %	82.35 %
Kappa value	2D	0.64	0.54	0.55
	2D+t	0.78	0.67	0.68

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript